

Review of “**Comprehensive analysis of molecular phylogeographic structure among the meadow jumping mice (*Zapus hudsonius*) reveals evolutionarily distinct subspecies**”

Authors: King et al.

Summary: The authors embark on a study to verify or refute the findings in Ramey et al. (2005) concerning the uniqueness of the federally endangered subspecies *Zapus hudsonius preblei*, the meadow jumping mouse. The authors use mtDNA data from two regions of the mitochondrion (control region and cytb) for a total of 1380 aligned bases and nuclear data from 21 microsatellite loci. Of the 12 described subspecies within the *Zapus hudsonius* species complex, the authors sample the geographically proximate five subspecies relative to the target subspecies of interest. They conclude that their data and associated analyses support the evolutionary distinctiveness of the subspecies in question, contrary to the results of Ramey et al. They then articulate why they believe they have the more robust result – based on better analyses of the data and better data. This study is, in general, well done with an excellent use of a broad range of statistical approaches to analyze data and good sampling at given locations for population genetic inference. However, there are also concerns with this study, as there are with the Ramey et al. study. The authors conclude that the differences are not merely differences between “lumpers and splitters”, but by my reading, this is indeed one of the many differences between the studies. So I will start this review by outlining those things done well and those things of concern by the King et al. study. I will then summarize my opinion with respect to this described subspecies based on the data from both studies.

Sampling

The most critical issue relative to the robustness of the conclusions drawn in this report is one of sampling. King et al. rightly point out that sampling is critical in intraspecific studies and is distinct from systemic studies. They critique Ramey et al. on a sampling strategy of few individuals per location with broad location sampling throughout known distributions. Instead, King et al. argue for dense sampling at specific locations with sparse sampling across locations throughout the distribution of the subspecies. King et al. correctly point out that the basis of inference by Ramey et al. (frequency differences instead of evolutionary relationships) is highly dependent upon sampling individuals at a given location with the Ramey et al. sampling design lacking in terms of individuals per site. Yet the conclusions reached by King et al. are also highly suspect in that leaving large geographic gaps between sampling sites when the taxon is known to range within those gaps leads to artificial inferences of population structure when, in fact, a gradient of variation may exist with gene flow across the gradient. Thus the optimal sampling strategy for such studies is often a combination of the two approaches, guided by preliminary examination of molecular genetic data (Morando et al. 2003). Both studies fail in terms of sampling strategy. The conclusions by King et al. of population structure are particularly highly suspect given the sampling design of their study. For example, King et al. have no *Z. h. preblei* from southern Wyoming, where according to the distribution map in Ramey et al., they are distributed and would be individuals most likely to show evidence of gene flow between *Z.h. campestris* and *Z. h. pallidus*, based on geographic proximity. Likewise, King et al. have no samples of *Z. h. pallidus* from western Nebraska, again

the region most likely to show signs of mingling of haplotypes. And again, King et al. have just a single locality sampled for *Z. h. luteus* and just two sites sampled for the critical *Z. h. campestris* and *Z. h. intermedius*. This is particularly problematic with the widespread distribution of *Z. h. intermedius* across 11 states with sampling in only the NE corner of South Dakota and an adjacent site in central Minnesota. Indeed, to do the study correctly, as King et al. point out, it would be ideal to sample the entire species complex throughout its range. Unfortunately, both studies fall short on this account as well. The central problem here is a taxonomic issue relative to the entire species complex, thus the entire species complex should be sampled to resolve the issue.

Data

Given the sampling caveats, the studies differ significantly in terms of the actual data used in analyses. Ramey et al. sampled morphological data which seems reasonable given the original designation was based on morphometric designations. King et al. ignore morphology all together. Both studies then incorporate microsatellite data and mtDNA data. The King et al. study is far superior in this respect as they have included nearly six times the number of microsatellite loci (overkill really) and an additional mtDNA locus (cytb). Importantly, King et al. correctly point out that phylogenetic inference is highly dependent upon the length of the sequence data used and the addition of the cytb sequence data coupled with a longer control region allows for more robust inference. King et al. also correctly point out that the cytb locus is a standard for species delimitation studies (as is COI – barcoding) and the control region has difficulties due to evolutionary constraints on this region. Thus the King et al. cytb data set is, in my opinion, far superior to the Ramey et al. CR data set for making phylogenetic inferences. Likewise, the 21 microsatellite loci are, by definition, better than only 5. Of course, 21 microsatellite loci are more than enough to determine family groups within a subspecies (Villanueva et al. 2002), let alone species/subspecies status. So in my mind, the microsat data are to some extent an overkill and fairly desperate attempt to find variation at all costs. The money would have been much better spent on sampling more samples from locations.

Criteria for Designation

A lesser but still significant issue is the designation of a criterion or criteria with which to test species or subspecies status. King et al. critique Ramey for employing criteria of reciprocal monophyly that are, they argue, too stringent for subspecies designation. In fact, Ramey et al. actually argue for exchangeability criteria which were designed precisely because the reciprocal monophyly criteria can not only be too restrictive but also relatively uninformative relative to adaptive variation. Ramey et al. use the frequency based approach (with inference limited due to the sampling difficulties outlined by King et al.) to test genetic exchangeability and use morphometric data to test ecological exchangeability. They fail to reject the null hypotheses of exchangeability on both accounts. King et al., on the other hand, adopt a criterion of population subdivision as equal to subspecific designation. They do not test within the exchangeability criteria at all, except that we might equate population subdivision as a measure of genetic exchangeability. Ecological exchangeability is ignored by King et al. Whether or not population subdivision is an appropriate criterion for subspecific designation is, unfortunately, in the eye of the beholder. This

is where we get to the lumpers versus the splitters. The central problem being that biologists have a hard enough time agreeing on what a species is, let alone a subspecies. King et al. appear to be arguing that any population genetic subdivision should be designated as a distinct subspecies and argue then that the clade of *Z. h. preblei/campestris/intermedius* should each be a distinct subspecies within that clade because they find evidence of population substructure. Ramey et al., on the other hand, argue that since they all form a clade, they are one subspecies relative to the other subspecies (see figure 7 of King et al. for the same result). The implication of the King et al. interpretation is that, given the relative levels of genetic differentiation, *Z. h. pallidus* and *Z. h. luteus* should be their own species. Yet King et al. seem to ignore these taxa and focus on the issue at hand. In looking at the King et al. criteria articulated, they are basically arguing for what Mortiz designated as “Management Units” as a subspecies designation and Ramey et al. are arguing for what Mortiz designated as “Evolutionarily Significant Units” as a subspecies. I believe both studies have diagnosed their respective units appropriately within these frameworks – given the caveats to both studies outlined here (sampling, data, analyses, etc.). They get different units because they are applying subspecies criteria at different levels of the evolutionary hierarchy (see Daniels et al. 2005 for further concerns with subspecies designations).

Analyses

The analyses done in this report are, in general, very well done. Indeed, the analyses are far superior in this report than those performed by Ramey et al. and, in my opinion, more appropriate for the data and the questions being addressed. I do, however, have a number of minor comments about the analyses and a few more significant questions/challenges for the analyses. I’ll start with the more major issues.

First, the central hypothesis that King et al. wish to test is one of panmixia versus structure. I’ve already discussed the difficulty in equating lack of panmixia with subspecies designation (indeed, this would be the most extreme definition of a subspecies ever used and seems to be employed by King et al.). However, the authors actually fail to even test panmixia. Furthermore, they also fail to provide basic measures of gene flow. I would suggest that King et al. use the recently described approach by Carstens et al. (2005) to test for population structure. This is an actual test for structure as opposed to testing for an F_{st} significantly different from zero and is most appropriate for the mtDNA data.

Second, there is clearly some structure within the *Z. h. preblei/intermedius/campestris* complex and according to the estimated network it corresponds nicely to the named subspecies. Now let’s ignore the sampling issues for a moment and take these results at face value. Then, as a conservation biologist, I would be very interested in the levels of gene flow among these subpopulations and the relative effective population sizes among these populations. Additionally, one can estimate both historical and current effective population sizes to gain insights into the current and past operators on population structure by comparing theta estimates of diversity with pi estimates (see Buhay and Crandall 2005 for details). Furthermore, one must ask about the relative timing of the substructuring and the cause of such historical divisions. While King et al. provide a network and argue effectively for its appropriateness relative to a phylogenetic tree (and certainly the NJ tree provided by Ramey et al.), they do not

take the analysis to its completion by performing a nested clade analysis on the data (Templeton 2004). Such an analysis would allow the partitioning of current from historical demographic and structuring events and allow for a comparison of the relative timing of divergence events. I suspect that what it would more likely do is point out the sampling wholes in the data, but it would pinpoint where additional samples should be collected. Thus such an analysis is a must. Additionally, King et al. make reference to timing of events in their summary arguments for the subspecies designations. On page 28, they discuss “apparently longstanding separation” and barriers to gene flow among subspecies that “have been of sufficient duration”, yet they never provide actual estimates of divergence times among these three subspecies and compare amongst the other two subspecies. The timing of the divergences is really critical too, because the gene frequency differences could simply be due to human activity partitioning up the habitat. If this species was historically widespread with reasonable gene flow and is now (recently) subdivided (as King et al. argue) and if this subdivision is a result of human impact, then the appropriate conservation action should actually be to **restore** gene flow among these isolated fragments to protect the genetic integrity of the species and restore historical processes that have been disrupted by human impact (see Crandall et al. 2000 for a discussion). So this becomes a really critical issue on the management implications of an inference of subdivision.

Another major issue with the analyses is that they are all done assuming the alternative hypothesis is true. The null hypothesis is panmixia or no differentiation among the subspecies. But much of the analyses are done assuming the current taxonomy is correct. This leads to real problems when you try to test this current taxonomy because you’ve already biased your results with it. This is an issue in the structure analysis in pooling data by subspecies name (page 12). Furthermore, King et al. do very little in terms of formal hypothesis testing relative to the subspecies designation. For example, with the mtDNA results, how different is a tree that puts a *Z. h. preblei* haplotype into the *Z. h. intermedius* clade? Can you reject that alternative. I suspect not. Likewise, in the structure analysis, you can test the next best alternative that breaks up the subspecies and see if you can reject it (using permutation testing). King et al. should use the evolutionary history to define the units to be tested and then test for significance of those against the taxonomic units.

Minor analysis issues

The ILD test is not a test for signal; it is a heterogeneity test.

ML search should be with random addition – not NJ starting tree; should not use “fast” option for bootstrapping; should perform Bayesian analyses

You don’t need to run Collapse. You can just run TCS with all your data and it will calculate the haplotypes for you and then also provide and incorporate frequency data.

Page 23 top, a little unfair as Ramey did use microsats and morphology (which was ignored here).

Page 24, King et al. report that there seven distinct clusters found with the microsat

data. Indeed, the initial finding was for three clusters, one of which was the *Z. h. preblei/intermedius/campestris* complex in question. It was only after further subdivisions based on the taxonomy (circular reasoning) that additional subdivisions were found. Assignment is not a good verification of the results (it is, in fact, simply the inverse of what the software has already partitioned). Can you reject one cluster? That's the question.

It would be better to have a few sequences per species and a few different species for your outgroups to see how robust the monophyly of the "species" is. Of course, that is sort of irrelevant given you haven't sampled all the subspecies within the species.

Summary Conclusion

If I were forced to draw a conclusion from this study, I would first of all include many caveats including all those listed above in terms of additional analyses, additional data collection, explicit hypothesis testing, and additional taxon sampling. Yet, given the results, my interpretation would be that figure 6 clearly shows three distinct clades that are also supported in Figure 7. These correspond to what I would call subspecies (given the current taxonomy and lack of sampling of the other subspecies), which include *Z. h. preblei/intermedius/campestris*, *Z. h. pallidus*, and *Z. h. luteus*. Indeed, these three subspecies also correspond to the three clusters found by structure. Thus, there seems to be agreement in both the nuclear and mtDNA in the rejection of genetic exchangeability between these three subspecies. Conservation implications should take into account these evolutionarily significant units. Imperative in appropriate conservation action is to determine if the structure found within the *Z. h. preblei/intermedius/campestris* is natural or human induced. This will dictate appropriate conservation action. This conclusion assumes that the data and analyses presented in these reports are reasonable and accurate (King et al. correctly point out the weaknesses of the Ramey et al. data with respect to museum samples etc.). Furthermore, this conclusion could very well be rejected with further and more appropriate sampling (as outlined above). Given the sampling issues, the prudent action at the moment is to leave the subspecies listed as is (endangered) and reevaluate after a full study of the species can be accomplished including extensive sampling across the distributional range of each subspecies.

References

- Buhay, J. E., K. A. Crandall. 2005. Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology* 14:4259-4273.
- Carstens, B. C., A. Bankhead III, P. Joyce, J. Sullivan. 2005. Testing population genetic structure using parametric bootstrapping and MIGRATE-N. *Genetica* 124:71-75.
- Daniels, S. R., N. J. L. Heideman, M. G. J. Hendricks, M. E. Mokone, and K. A. Crandall. 2005. Unraveling evolutionary lineages in the limbless fossorial skink genus *Acontias* (Sauria: Scincidae): are subspecies equivalent systematic units? *Molecular Phylogenetics and Evolution* 34:645-654.
- Morando, M., L. J. Avila, and J.W. Sites, Jr. 2003. Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongates-kriegi*

- Complex (Squamata: Liolaemidae) in Andean-Patagonian South America.
Systematic Biology 52(2):159-185.
- Templeton, A. R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Molecular Ecology 13:789-809.
- Villanueva, B., E. Verspoor, P. M. Visscher. 2002. Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. Animal Genetics 33(1):33-41.